Hello "Stakeholder",

In regards to the Fetch Rewards data that was sent by your team, we have been able to review the data files, and we have some questions, feedback and results.

Question:
- All the date formats in the files are of numeric integer value. What is the conversion logic to reverse them to date/time format?
    - We were able to apply our own logic, from our own assumptions, but would like confirmation that the logic we applied is correct.

Feedback:
- Users data file has duplicate entries. We were able to extract only unique users (and not the duplicates).
- There are a handful of duplicated BarCode entries (7) in the Brands file. Even though there are not a lot of duplicates, it can definitely yield wrong results during analysis.
- There are 148 Receipts that do not have an associated User from the Users file. This can also yield wrong results on those queries/requests that involve User related transactions.
- And most importantly, 98% of the Receipt Items extracted from the Receipts file, do not have a valid BarCode. This presents a major issue in generating queries/requests that involve Brands and Receipts,since there is no way to associate the 2 data sources.

Results
- On the other hand,we were able to satisfy one of the requests.

    Question : "When considering *average spend* from receipts with 'rewardsReceiptStatus' of 'Accepted' or 'Rejected', which is greater?"

    Answer : "Rejected" Status is greater (23.33) vs "Accepted" (0.0)

    "Accepted" status does not have any entries in the Receipts data file.

Average spend
from receipts
with
'rewardsReceiptS
of 'Accepted' or
'Rejected'

F11
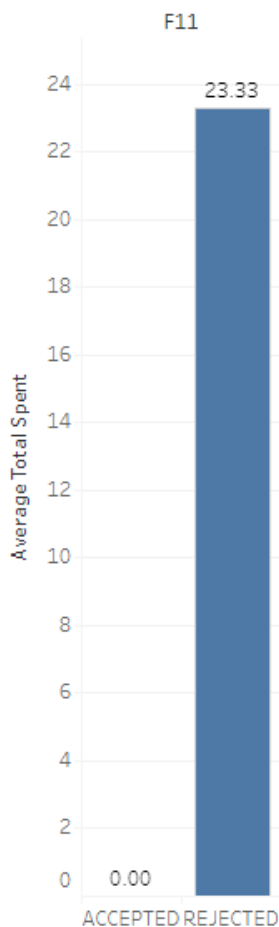
In order to satisfy the remaining requests, we would like to discuss the data quality issues with your team, and also provide you with our data quality requirements.

Some of these requirement would be, for example:
- Cross validating the Receipts data source to make sure all Brands exist.
- Validating that each Brand has a unique and valid BarCode
- Making sure that all entries are unique in each data file.

By not having the Data Quality measures in place, this can potentially yield the wrong result in our Production environment, since any analysis outcome regarding Brands and Receipts can be misleading and incorrect.

Hope to hear from you soon
Regards,

-Esteban Ordenes
Data Analyst Team